DOCUMENT RESUME

ED 292 831 TM 011 209

AUTHOR Wall, Terry Cobb
TITLE The Illusion of Independence in Evaluation.
PUB DATE [84]
NOTE 13p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; *Evaluation Methods; Evaluation Problems; Evaluators; Program Evaluation

ABSTRACT

Values and beliefs underlying assumptions concerning the independent evaluations of persons and programs involved in the educational process are discussed. Particular emphasis is placed on M. Scriven's treatise on bias in evaluation (1976). Works by E. R. House, R. E. Stake, L. Stenhouse, M. Parlett and D. Hamilton, and D. L. Stufflebeam are also considered. Independence in evaluation is valued on the basis of shielding the observer from sources of bias and enhancing the observer's credibility as an evaluator. Issues of objectivity and intersubjectivity are central to any analysis of the enterprise of evaluation. Other issues of concern include assumptions underlying the empirical-analytic paradigm, behavioral measures, differences distinguishing program evaluation and program development, experimental and psychometric traditions, and illuminative evaluation. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED292831

THE ILLUSION OF INDEPENDENCE IN EVALUATION

Terry Cobb Wall

North Carolina State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Terry Cobb Wall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 209

THE ILLUSION OF INDEPENDENCE IN EVALUATION

Terry Cobb Wall
North Carolina State University

The independent evaluation is something of an institution among educators and developers of social programs in general. The goal, if not the practice, of valuing justified criticism is given lip service in most professional circles. The critical question is what it takes to persuade interested parties that they should place their faith in a given evaluation. There are assumptions which underlie the concept which is expressed when the words "independent evaluation" are spoken. Imbedded in those assumptions are values and beliefs, the examination of which is the aim of this paper.

"Since it is a universal truth that self-evaluation is unlikely to be reliable, it is a necessary consequence of interest in truth that one supplement self-evaluation," writes Michael Scriven in a useful treatise on bias in evaluation (1976). "Hence anyone interested in improving his or her own performance must arrange for or endorse some kind of independent evaluation." In this case, the context which provides the background for the use of the word "independent" suggests a meaning related to an evaluator's distance from biasing influences. But there is a limit to the distance from which a program may be observed without destroying the observer's chances of producing a report which has any relevance for practitioners.

Scriven's examination of evaluation bias makes the point

that "the quest for objectivity via the criterion of independence often leads to the use of 'external' evaluators" (p. 128). In this instance, independence refers to issues that would provide an evaluator with more or less credibility. Staff evaluators, for instance, enjoy little credibility because they do not bring to the evaluation task an outsider's fresh perspective.

Independence is valued, it appears, for what it can do to achieve two things which are important to evaluation studies. First, the observer must be shielded from sources of bias. In addition, independence from involvement in a program enhances the observer's credibility as an evaluator. Can a claim be supported that the gains which result from the cultivation of "independence" in evaluation are of sufficient magnitude to justify the precautions taken? Or have concerns about bias and credibility constrained evaluative inquiry beyond what is reasonable? In other words, does concern with independence result in a net gain or loss of power for evaluation methodology?

Scriven observes, without taking a stand on the propriety of common practice, that the skills of outside evaluators are often desired by professionals in education. He credits such arrangements for what they achieve, i.e., facing the opportunity for bias and acting to control and minimize it. But, in the final analysis, he finds fault with the prevalent understanding of the principle by which most educational evaluations are guided, i.e., objectivity.

The crux of the confusion over criteria for acceptable evaluation, apparently, lies in misconceptions about objectivity. Arguments attributed to Scriven are reviewed in a chapter by

House on the logic of evaluative argument (House, 1980, pp. 86-88). The theme of most definitions of objectivity is that there is something outside the mind that is verifiable through public or intersubjective agreement and that one can express or prove such things without influence from personal feelings. An evaluation which can do so is said to be objective.

Scriven has isolated one meaning, as just explained, and ascribes that understanding to the quantitative sense in which the term "objectivity" is used. At this point, House's summary of an article by Scriven (1972) is useful:

> In the quantitative sense of the term, one person's opinion about something is regarded as being subjective--the disposition of the individual. The common experiencing makes the observation public through intersubjective agreement. . . . The qualitative sense of objectivity is quite different. . . . Being objective means that the observation is factual, while being subjective means that the observation is biased in some way.
>
> House (1980), p.86

So an observation can be quantitatively subjective (one man's opinion) and also qualitatively objective (actually unbiased and true). Measurement on a quantitative scale by mechanistic means becomes the indicator of truth because the interjudge reliablity is higher than if each judge depended on human perception instead.

Again, House's summary of Scriven is instructive. "From this idea--that what cannot be directly experienced by others cannot be taken seriously as science (intersubjectivism)--has developed the concept of objectivity as the externalization of all references so that multiple witnessing can be achieved..." Claims which can be verified through external or objective means

permit replication. Replication is the hallmark of reliability, but can do nothing to solve the problem of bias. Highly reliable instruments may be of questionable validity.

Scriven's history lesson tells us that two separate modes of inquiry, one quantitative and one qualitative, have become confused and have resulted in a mistaken notion that repeatability constitutes proof of objective truth. House restates Scriven's tenets (1980, p. 86):

> 1--confused thinking has led to this indefensible position (that agreement on externalized measures constitutes truth);
>
> 2--not only is intersubjective verification not a guarantee of truth, it is not necessary;
>
> 3--truth is an ideal which can only be approximated through an interplay of introspection and public verification.

The subjective element, then, is neither undesirable nor avoidable when it comes to evaluation. Furthermore, since there are only limited available "checks" on an evaluator's freedom from bias, there are no guarantees of truth. Tendency statements, analogies, approximations--these are the knowledge claims available in the social sciences. Evaluation is one of the fields which is based on just this kind of "weak knowledge," Scriven concludes. The best way to assure quality in evaluation, he tells us, is to make sure the evaluators get evaluated.

Scriven has argued that there is simply not a good match between criteria for good evaluation and the assumptions which underlie the empirical-analytic paradigm. Social science inquiry fails to have the predictive power or to establish the linear relationships which are the stuff of which one generally assumes that "science" is made. But in exchange for the lack of predictive

and explanatory power, there is a gain in both richness of description and opportunities for discretionary judgments by the observer.

Stake has thoroughly outlined the distinguishing characteristics of evaluation research, referring to the multiple measures and data sources as "the full countenance of educational evaluation" (Stake, 1967). Far from advocating the temperance of expectations of the highly subjective enterprise called evaluation, Stake celebrates its potential for capturing the complex, dynamic nature of education.

By attending to much more than student behaviors as measures of outcome, the evaluator increases the chances that a study will uncover durable relationships which may exist between and among outcomes and/or other observables. Some of the many variables include: intent of the program designer, antecedent conditions inherent in the program's contextual setting, and a full range of teaching and learning transactions.

Stake's model is one of several refinements of evaluation methodology, with the creators of which Stenhouse (1984) feels a certain kinship. Approving of their attempts at innovation, he writes that there are "new wave evaluators who, by destroying the notion of an objective assessment . . . have created methods [in which] there is no place for the independent evaluator. The evaluative element is an integral part of . . . research design, and you cannot be a researcher without being an evaluator. The evaluator should be made responsible to--or, if you like, should be--the project director" (p. 85).

Stenhouse uses the term "independent" in a way that reveals

his distaste for clear distinctions among roles. A good practitioner is always evaluating, as the good researcher is always in touch with practice and its need for improvement. Independence is to be avoided, since it suggests actors out of touch with each other in an endeavor designed to operate as an organic whole.

It is possible that Stenhouse misrepresents the current state of evaluation methodology when he makes the claim that distinctions between program development and program evaluation have begun to disappear. But there is no denying that he shares a perspective with more than a few contemporary writers on evaluation that "some of the classic ideas about evaluation are fundamentally flawed," (Stufflebeam, 1984). One of those ideas, from Stenhouse's point of view, is that truth exists "out there" somewhere. "The illusion of the independent evaluation is the illusion of positivist social science that there can be some independent and disengaged view of truth," Stenhouse writes.

Parlett and Hamilton (1976), further examples of "new wave" voices, attribute the present methodological and conceptual disarray in the field of evaluation studies to a lack of coherent or agreed upon frames of reference. For too long, they say, the new field of evaluation was inappropriately dependent on an agricultural-botany paradigm which artificially restricted the scope of studies and diverted attention from questions of interest to practitioners.

A method which Parlett and Hamilton call "illuminative evaluation" combines data collection techniques to facilitate the cross-checking of what must inevitably be otherwise tentative

findings. The goal of illuminative evaluation is not truth, or even unbiased measures of effectiveness, but "attempts to explain" programs as they operate in their peculiar contexts. Tacit assumptions and hidden curricula, sometimes perceptible in language conventions, are important areas for evaluative exploration, they say.

The subjective element is present with open-ended techniques. It is inevitable. When the investigator abandons the agricultural-botany paradigm, his role is necessarily redefined to include responsible use of his primary instruments--interpretive human insight and interpersonal skills. Parlett and Hamilton encourage evaluators to make discretionary and potentially biasing judgments. In so doing, they place a premium on subjective and personal interpretation. Forms of research do not exist, they believe, which are immune to prejudice, experimenter bias, and human error.

How might "independence" in evaluation be seen from this point of view? The illuminative evaluator's goal is not to arrive at independent, replicable decisions. In fact, the investigator does not arrive at decisions at all. Illuminative evaluation concentrates on the information-gathering component of evaluation, in order to contribute to decision-making. Ideally, evaluation output will be regarded as useful, intelligible and revealing by those involved with the enterprise, whose job it is to make decisions based on the investigator's portrayal of their complex reality.

Experimental and psychometric traditions have dominated evaluation research, but the literature reflects the steadily

increasing influence of social anthropology and participant observation research in sociology. Evaluators who follow the latter traditions see programs as inseparable from the surroundings in which they are introduced and developed. They search for multiple effects of programs, as opposed to single output measures.

Another distinctive innovation in evaluation is the insistence that evaluation serve the information needs of decision-makers and avoid any tendency to focus on the more centralized, bureaucratic concerns. Such a position has been proposed by Stufflebeam (1969; 1984) and is shared by a growing number of evaluators. As a consequence, there is interest in approaching each evaluative study by creating a broad data base, reducing emphasis on test scores, and including increasing numbers of people in evaluations as data sources and as audiences.

House (1976) identifies this movement as the pluralist-intuitionist "counterstream" in educational evaluation. The identifying characteristic of this school of thought is the interest in what he calls democratizing knowledge demands, the implicit goal of which is to encourage participation, similar to the idea behind pluralist politics. In radical ways, such evaluators depart from the utilitarian ethics, which has for so long mistaken impersonality for impartiality.

The utilitarian practice of mediating among special interests requires an evaluator to remain detached and truly disinterested. What people really desire, according to House (1980, p.187), is "an impartial evaluation, one in which group

interests are represented, in which groups are treated fairly, and in which the evaluator is not excessively promoting his own interests."

The evaluator who presents information for collective determination of choice strives for a study in which all relevant interests are represented (rather than none). A distinterested, detached evaluation has the anemic characteristic of being irrelevant since it does not represent the interests of those most concerned, House reasons. Neither disinterest nor objectivity is sufficient as the basis for an evaluation so conceived. Instead, there is a requirement that evaluation promote the understanding of many viewpoints (House, 1980, p.191).

Stufflebeam (1984) argues that the most important purpose of evaluation is not to prove but to improve. One option for evaluators is to focus on informing multiple audiences for the purpose of their collective deliberation with regard to the elimination and development of programs. In so doing, evaluators will have radically redefined their own roles, with the result that the single criterion of "independence" may no longer be central to the question of the quality and overall usefulness of the evaluative study.

Evaluation, newly conceived, may incorporate elements of inquiry, valuing, and social change, and so can be seen to share common attributes with the three major areas of human enterprise: sciences, the arts, and the world of practical affairs (Smith, 1981, p.30). Evaluative inquiry will seek to know what is true; the aspect of valuing will pursue the beautiful and good; and the

element of social change will lead evaluators to ask what is viable and fair as humankind seeks to improve the human condition.

I close with a bit of philosophy which seems consistent with an expanded and more complex understanding of education evaluation. It is drawn from an essay on post-modernism by the anthropologist Frederick Turner:

> We are at a curious juncture in the history of science and technology. The empiricism of the Renaissance gradually flattened out the ancient hierarchy of the universe and broke up the Great Chain of Being.... Not that the empirical detour was unnecessary.... we needed three centuries of tearing things apart to see how they worked, to be able to come back to a coherent world, this time with the powers and knowledge we always felt were our birthright--powers and knowledge we had mimed with magic.

REFERENCES

House, E. R. (1980) Evaluating with Validity. Beverly Hills: Sage Publications.

House, E. R. (1976) "Justice in evaluation," in G. V. Glass (ed.) Evaluation Studies Review Annual, Vol. 1. Beverly Hills: Sage Publications.

Parlett, M. & Hamilton, D. (1976) "Evaluation as illumination: A new approach to the study of innovative programs," in G. V. Glass (ed.) Evaluation Studies Review Annual, Vol. 1. Beverly Hills: Sage Publications.

Scriven, M. (1976) "Evaluation bias and its control," in G. V. Glass (ed.) Evaluation Studies Review Annual, Vol. 1. Beverly Hills: Sage Publications.

Scriven, M. (1972) "Objectivity and subjectivity in educational research," in L. G. Thomas (ed.) Philosophical Redirection of Educational Research. National Society for the Study of Education (Australia).

Smith, N. L. (1981) "Developing evaluation methods," in N.L. Smith (ed.) Metaphors for Evaluation. Beverly Hills: Sage Publications.

Stake, R. E. (1967) "The countenance of educational evaluation," in B. R. Worthen & J. R. Sanders (eds.) Educational Evaluation: Theory and Practice. Worthington, OH: Wadsworth.

Stenhouse, L. (1984) "Evaluating curriculum evaluation," in Clem Adelman (ed.) The Politics and Ethics of Evaluation. London: Croom Helm.

Stufflebeam, D. L. (1984) "Evaluation as strategy for the pursuit of excellence." Kalamazoo: College of Education, Western Michigan University. (mimeo)

Stufflebeam, D. L. (1969) "Evaluation as enlightenment for decision-making," in Beatty (ed.) Improving Educational Assessment. Washington, D.C.: ASCD, NEA.

Turner, F. (1984) "Escape from modernism," Harper's Magazine, November: 47-55.